Subband Coding Methods for Seismic Data Compression*

Aaron Kiely and Fabrizio Pollara
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109
e-mail: aaron@shannon.jpl.nasa.gov, fabrizio@shannon.jpl.nasa.gov

Abstract

This paper presents a study of seismic data compression techniques and a compression algorithm based on subband coding. The algorithm includes three stages: a decorrelation stage, a quantization stage that introduces a controlled amount of distortion to allow for high compression ratios, and a lossless entropy coding stage based on a simple but efficient arithmetic coding method. Subband coding methods are particularly suited to the decorrelation of non-stationary processes such as seismic events. Adaptivity to the non-stationary behavior of the waveform is achieved by dividing the data into separate blocks which are encoded separately with an adaptive arithmetic encoder. This is done with high efficiency due to the low overhead introduced by the arithmetic encoder in specifying its parameters. The technique could be used as a progressive transmission system, where successive refinements of the data can be requested by the user. This allows seismologists to first examine a coarse version of waveforms with minimal usage of the channel and then decide where refinements are required. Rate-distortion performance results are presented and comparisons are made with two block transform methods.

1 Introduction

A typical seismic analysis scenario involves collection of data by an array of seismometers, transmission over a channel offering limited data rate, and storage of data for analysis. Seismic data analysis is performed for monitoring earthquakes and for planetary exploration as in the planned study of seismic events on Mars. Seismic data compression systems are required to cope with the transmission of vast amounts of data over constrained channels and must be able to accurately reproduce both low energy seismic signals and occasional high energy seismic events.

We describe a compression algorithm that includes three stages: a decorrelation stage based on subband coding, a uniform quantization stage, and a lossless entropy coding stage

*The research described in this article was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

based on arithmetic coding. Rate-distortion performance results are presented and comparisons are made with two block transform methods: the Discrete Cosine Transform (DCT) and the Walsh-Hadamard Transform (WHT).

Subband coding methods are particularly suited to the decorrelation of non-stationary processes such as seismic events. For most seismic data, signal energy is more concentrated in the low-frequency subbands, which suggests the use of nonuniform subband decomposition. The decorrelation stage is implemented by quadrature mirror filters using a lattice structure. Adaptivity to the non-stationary behavior of the waveform is achieved by dividing the data into blocks which are separately encoded.

The compression technique described in this paper can be used as a progressive transmission system, where successive refinements of the data can be requested by the user. This allows reconstruction of a low resolution version of the waveform after receiving only a small portion of the compressed data. This could allow seismologists to make a preliminary examination of the waveform with minimal usage of the channel and then decide where high resolution refinements are desired.

In general, given a fixed transmission rate, lossy compression algorithms applied to high accuracy instruments deliver higher scientific content than lossless compression methods applied to lower accuracy instruments.

2 Subband Decomposition

In the *analysis stage* of subband coding, a signal is filtered to produce a set of subband components, each having smaller bandwidth than the original signal. Because of this limited bandwidth, each component is downsampled, so that the subband transformed data contains as many data points as the original signal. The subband components are then quantized and compressed. In the *synthesis stage*, the reconstructed signal is formed by adding together the subbands obtained by applying the inverse filters to upsampled versions of the subband components.

The analysis and synthesis filters used here are FIR quadrature mirror filters (QMF) implemented using the lattice structures shown in Figures 1 and 2 which are described in [6, 1]. Analysis and synthesis quadrature mirror filters of order $2M$ are implemented using an M stage lattice structure. Suitable lattice filters can be found in [1, p. 267], [6, p. 310].

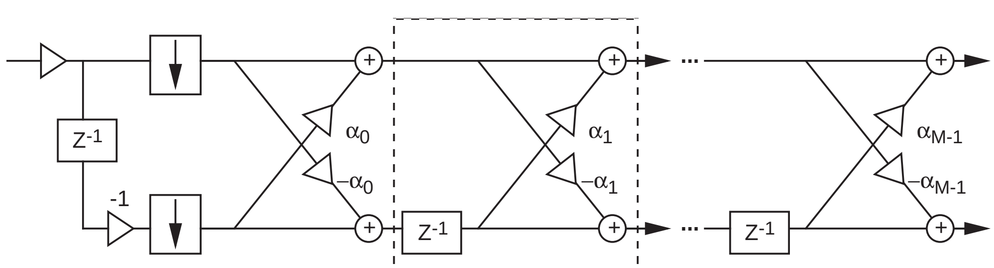

Figure 1: Analysis filter structure. (The stage inside the box is repeated.)

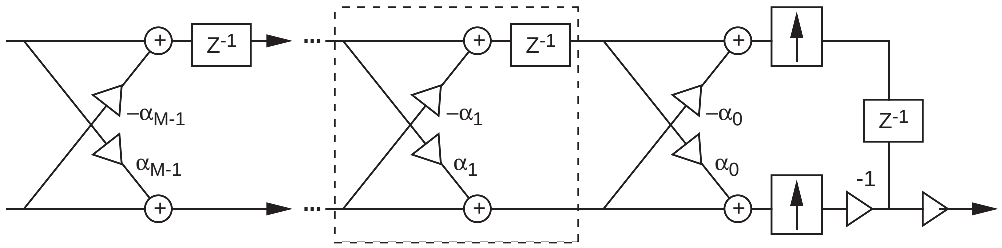

Figure 2: Synthesis filter structure.

For most seismic data samples, signal energy is concentrated primarily in the low subbands[1]. The uneven distribution of spectral energy in seismic signals provides the basis for subband coding source compression techniques. For effective signal coding, subspectra containing more energy deserve higher priority for further processing.

A subband decomposition that tends to work well for seismic data is the "dyadic tree" decomposition shown in Fig. 3. The signal is first split into low- and high-frequency components in the first level. A two-band subband decomposition uses high-pass and low-pass digital filters to decompose a data sequence into high (H) and low (L) subbands, each containing half as many points as the original sequence. The filter is repeated to further decompose the low subband. This process may be repeated several levels.

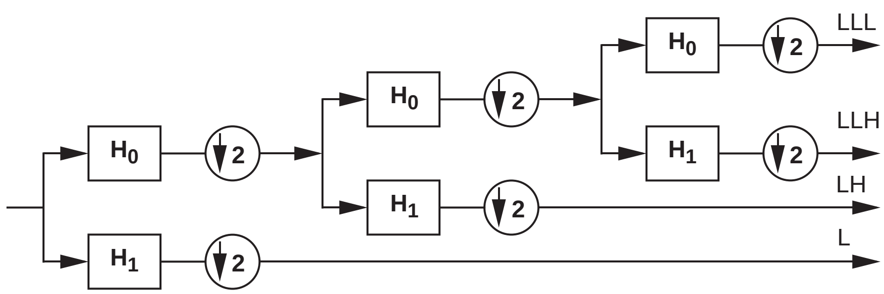

Figure 3: Subband decompositions.

Increasing the number of subbands produces diminishing rate-distortion returns, with gains often observable only at very high compression ratios. One reason for this is that after several decompositions, the energy is no longer so highly concentrated in the lowest subband.

So that a filtered block has the same length as the original, each block is periodically extended (i.e., repeated in time) before filtering, and the components corresponding to a single period of the filtered extended signal are taken as the filtered signal. If this operation were not performed, the length of the filtered signal would exceed the original block length. An unfortunate side effect of periodic extension is that it often produces high frequency

[1]This generally applies to the "EHZ" and "BHZ" seismic data components, which have sample rates of 100 and 20 samples per second respectively. Energy in "LHZ" data, which has sample rate of only one sample per second, is typically not as concentrated in the low frequencies. However, because of the much lower sample rate, compression of this component is not as important as the others. A different subband decomposition could be implemented to accommodate this type of data.

components at the edges of data blocks, an effect whose impact increases with filter length. These components are not as easily compressed as the rest of the subband data, and are separated for compression purposes. Longer filters are also more likely to introduce noticeable spurious effects at the onset of a high energy seismic event, as we shall see in Section 6. It is also worth noting that longer filters generally do not dramatically outperform shorter filters, as we will see in the following section.

3 Comparing Subband Coding to Block Transforms

For comparison purposes, we also examined the discrete cosine transform (DCT), a popular technique used in the compression of two dimensional data (e.g., images). A general description of the DCT as used in the JPEG compression algorithm can be found in [4, pp. 113-128]. The DCT can also be applied to one dimensional data, as is done here.

The data are partitioned into blocks of length 8, the DCT of each block is computed using the 8×8 DCT matrix, and these transformed values are uniformly quantized. A different quantizer stepsize could be used for each coefficient, but in practice, for most seismic data samples, near optimum performance is obtained when all quantizers use the same step size. The quantized coefficients are arranged in groups of eight blocks for subsequent coding, so that 64 transformed coefficients are encoded at a time. In this way the procedure is similar to a one dimensional version of the JPEG algorithm. The lowest frequency (DC) quantized coefficients are encoded using DPCM and Huffman coding, except at very low rates, when a runlength code is used. The remaining (AC) coefficients are runlength encoded, in order of increasing frequency. The runlength encoding used is the same as that described in [4, pp. 114-115].

We also used the same algorithm with an 8×8 Walsh-Hadamard transform (WHT) in place of the DCT, separately encoding each coefficient. The WHT performed uniformly worse (see Figure 4). To make a fair comparison with subband coding, we compared the block transform compression methods to subband coding combined with Huffman coding of the quantizer output, rather than the arithmetic coding procedure to be described in the next section.

Rate-distortion curves for a seismic data sample using these different techniques are shown in Fig. 4. The labels on the curves corresponding to subband coding identify the number of subbands and the particular filters used. For example, "3B8L" refers to a three band decomposition using an order 8 FIR filter. In terms of RMSE, subband coding is able to outperform the DCT and WHT with only moderate complexity.

4 Entropy Coding Stage: Arithmetic Coding

Anyone who has experienced an earthquake knows that the energy present in a seismic signal can vary tremendously over time. Consequently, seismometers have a large dynamic range, and it is desirable to have an adaptive compression system capable of transmitting low energy and high energy signals reliably. In a progressive transmission system, such as the one outlined here, each successive data segment transmitted provides higher resolution

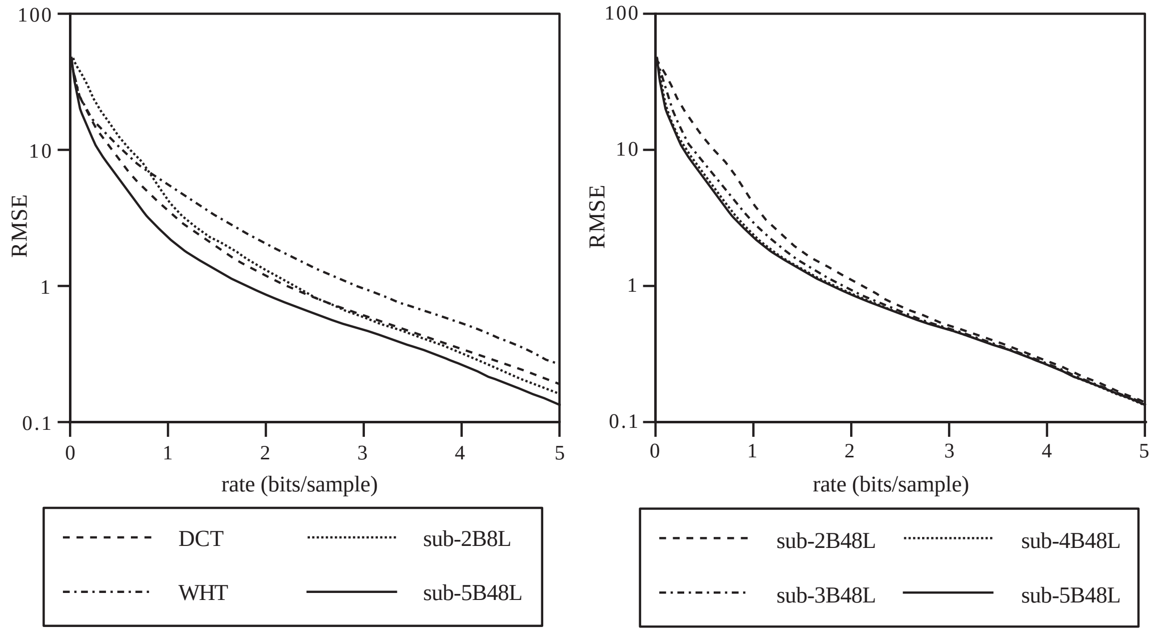

Figure 4: Rate-distortion performance for various compression techniques applied to a seismic data sample.

information about the signal, until the allocated rate is exhausted. In this manner, the resolution automatically increases for more compressible signals.

A block of m data samples produces m subband coded samples. Because of the down-sampling operation, half of these are high-subband samples, one fourth are low-high-subband samples, etc. All of the samples from a particular subband are quantized and encoded together block-adaptively. Because this is a block-to-block encoding procedure, the effects of a channel error are confined to the block during which that error occurs. The block encoding provides the additional benefit of adaptivity.

The output of the subband coding stage is a sequence of real numbers that are quantized and then compressed. For seismic data, as with many other types of data, these components are generally zero-mean, roughly symmetric, and have probability density that is decreasing as we move away from the origin. This is illustrated in figure 5, which gives an empirical probability density function (pdf) of signal amplitude from a low-pass filtered seismic data sample.

The compression scheme we use is bit-wise arithmetic coding [2]. A high resolution quantizer is used, and the quantized values are mapped into fixed-length binary codewords. Figure 6 illustrates the bit assignment for a four bit quantizer: the first bit indicates the sign of the quantizer reconstruction point, and each successive bit gives progressively higher resolution information. Because the pdf is zero mean and decreasing as we move away from the origin, a zero will be more likely than a one in every bit position. This redundancy is exploited using a binary arithmetic encoder to achieve compression.

Codewords corresponding to each subband are grouped together. The sign bits of the codeword sequence are encoded using a block-adaptive binary-input binary-output arithmetic encoder described in [2]. The next most significant bits are similarly encoded, and so

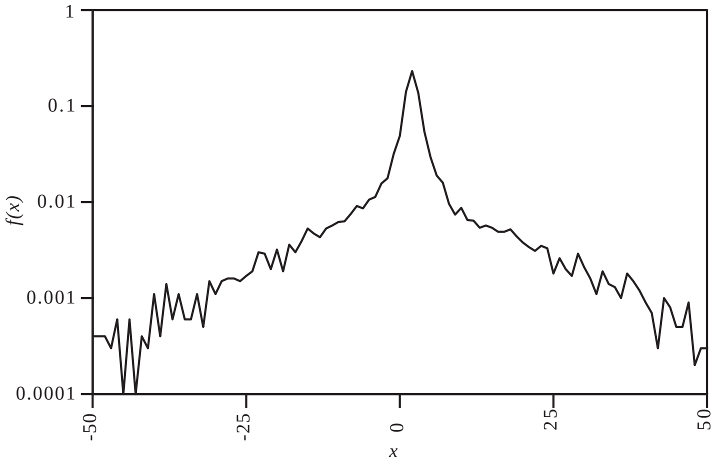

Figure 5: Empirical pdf for low pass subband filtered data.

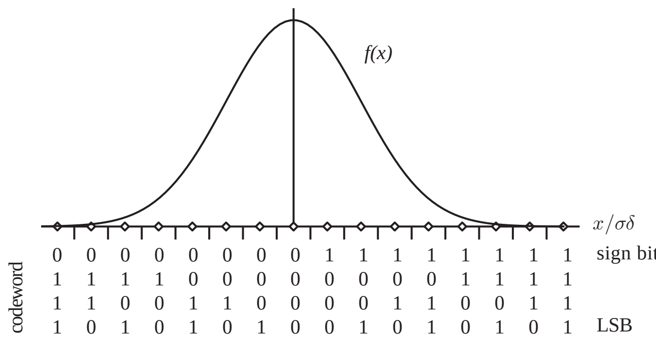

Figure 6: Codeword assignment for the four bit quantizer.

on. Each bit sequence (or layer) is encoded independently— at the ith stage the arithmetic coder calculates (approximately) the *unconditional* probability that the ith codeword bit is a zero.

The obvious loss is that we lose the benefit of inter-bit dependency. E.g., the probability that the second bit is a zero is not in general independent of the value of the first bit, though the encoding procedure acts as if it were. Traditional Huffman coding of the quantized samples does not suffer from this loss. However, for many sources, such as Gaussian and Laplacian sources this loss is quite small [2]. In fact, for many practical sources with low entropy, this technique has lower redundancy than Huffman coding, because the arithmetic coder is not required to produce an output symbol for every input symbol.

Because the inter-bit dependencies are ignored, very little overhead information is required (i.e., long tables of Huffman codewords are unnecessary). The overhead required for bit-wise arithmetic encoding increases linearly in the number of codeword bits. By contrast, the overhead of block-adaptive Huffman coding increases exponentially in the number of codeword bits unless we are able to cleverly exploit additional information about the source [3].

Another advantage is that, as we'll see in the next section, this technique is naturally progressive and gives a simple means of handling situations where we are rate constrained:

we continue encoding the codeword bits corresponding to higher levels of detail until the allocated rate is exhausted. The distortion is automatically reduced for "more compressible" signal blocks (e.g., low energy seismic waveforms) — when the most significant bits can be efficiently encoded, we are able to send additional (less significant) bits, so the encoder resolution increases automatically. This would mean, for example, that a block having 6 bit resolution might be followed by a block having 8 bit resolution.

5 Progressive Transmission Behavior

In designing a compression system to be used in progressive transmission or in situations where rate constraints may result in the loss of data, it is important to consider the rate-distortion behavior of the system when only portions of the compressed data have been received. Such performance can be improved simply by careful choice of the order in which the compressed data are transmitted.

The typical characteristics of subband filtered seismic data motivate our transmission strategy. Because the probability density for subband filtered seismic data is generally zero mean (see Figure 5), the sign bit layers of each subband usually have high entropy. Because the energy in seismic waveforms is often quite small, the high order bit layers (excluding the sign bit) often consist entirely of zeros or can be readily compressed using the block-adaptive arithmetic encoder. Finally, as mentioned in Section 2, periodic extension of the data is required in the subband filtering stage, which often produces high frequency components at the start of data segments. These initial values, which we call transients, are encoded separately from the rest of the data. All but the lowest subband contain these transients.

Generally speaking, we transmit compressed data ordered from most significant bit layer to least significant layer, and within this order, proceeding from lowest frequency to highest frequency subband. Initially, we skip the sign bit layer and begin with the next most significant bit layer. If this layer consists entirely of zeros, (which is usually the case), a single "0" is transmitted and we move on to the same layer in the next higher subband. For every subband, a "0" is transmitted for each layer consisting entirely of zeros until a "1" is transmitted at some layer ℓ, denoting that the ℓ^{th} layer is not all zeros. At this point, we transmit the sign bits (using the block-adaptive arithmetic coding procedure already described). Then the transients for the subband are transmitted using runlength encoding of the leading zeros, and then the (compressed) ℓ^{th} bit layer is transmitted. Then we proceed to the ℓ^{th} layer for the next higher subband. Each subsequent bit layer of the subband is sent, compressed by arithmetic coding.

Because the order of transmission is determined using a rather simple decision procedure, the additional overhead required to describe the transmission order is quite small — it consists only of occasional one bit flags.

The rate-distortion progressive transmission performance of this system for one seismic data sample can be seen in Figure 7. The highest rate point of each curve is the final design goal, and rest of the curve shows the rate distortion performance when the signal is reconstructed using only portions of the data. It is remarkable that the curves are nearly indistinguishable. Note that a system designed to transmit at a rate of 5 bits per sample (bps) but cut off at only 2.5 bps performs almost as well as a system designed to operate at

2.5 bps.

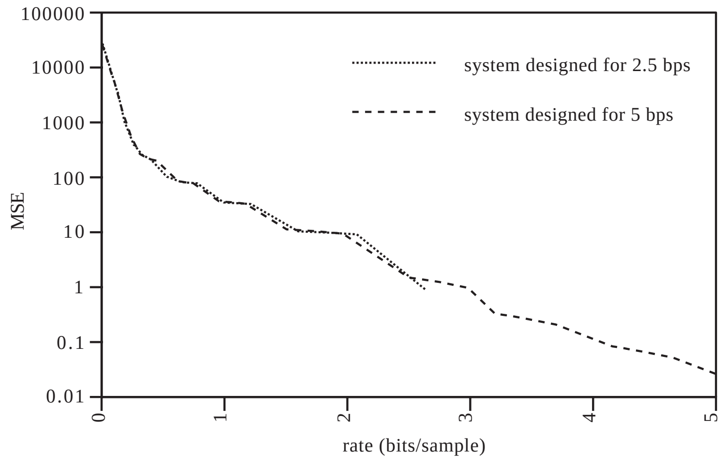

Figure 7: Progressive transmission performance.

6 Distortion Measures and Artifacts

In the previous sections, we have been mostly concerned with the mean square error distortion measure. However, mean square distortion may not be a sufficient indicator of fidelity for seismic analysis purposes. For example, Spanias et al. [5] examined the effect of transform data compression methods on estimation of the body wave magnitude, which they call "the key parameter used in seismic analysis." Other distortion measures may be more relevant, depending on the interests of the seismologists who will ultimately analyze the data. Unfortunately, we do not know of a distortion measure which seismologists will widely accept as the most useful.

Artifacts are erroneous features that may appear in the reconstructed waveform. Different algorithms create different artifacts depending on their mode of operation. For example, "blockiness" is an artifact commonly associated with block transforms such as the DCT, while "ringing" may be produced by subband coding using a filter with too sharp a response. Even a given algorithm may exhibit different artifacts depending on the bit rate at which it is operated. Some artifacts may be more objectionable than others for correct waveform interpretation.

In this section we illustrate two artifacts that may be observable in subband coding depending on the mode of operation and the compression ratio. Understanding the causes and cures for such artifacts allow seismologists to give meaningful feedback to engineers in deciding what features of a compression system are most important.

We are actively trying to engage the seismology community to characterize any essential artifacts produced by the proposed method [7]. One of the results of this interaction was the objection of seismologists to the precursor artifact created by a particular subband filter, as shown in Fig. 8(b). After determining that such an artifact was due to a filter with too sharp a response, we experimented with different, shorter filters producing the result shown in Fig. 8(c), which reduces the precursor problem while preserving essentially the same compression ratio.

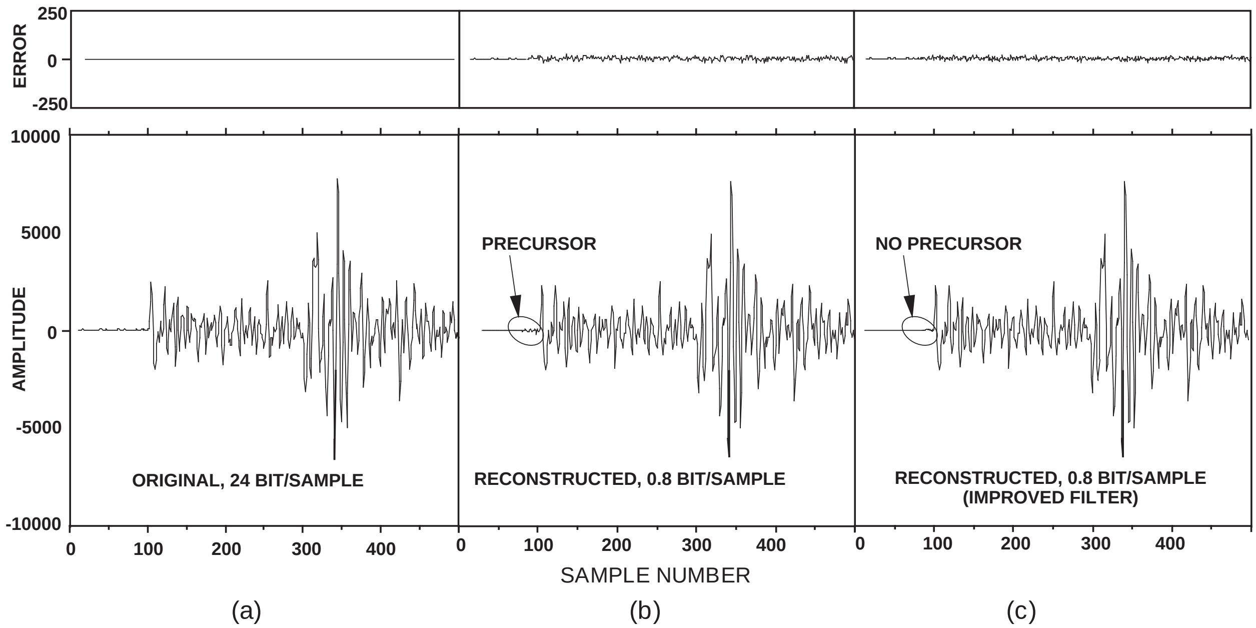

Figure 8: Original and reconstructed waveforms for two different filters.

A different artifact is introduced when the quantizer stepsize is quite large (this equivalent effect may occur if the waveform is reconstructed using only a portion of the data). In this case, each subband will have low resolution, and because most of the energy is contained in the low frequencies, the high frequency subbands may all be zeroed out. This produces the interesting smoothing effect that can be observed in the periodogram of the reconstructed waveform shown in Figure 9. If this frequency range has more significance than the others, the corresponding subbands could be assigned higher priority in the transmission and quantization stages.

References

[1] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Processing*, San Diego California: Academic Press, 1992.

[2] A. B. Kiely, "Bit-Wise Arithmetic Coding for Data Compression," *The Telecommunications and Data Acquisition Progress Report 42-117, vol. January-March 1994*, Jet Propulsion Laboratory, Pasadena, California, pp. 145-160, May 15, 1994.

[3] R. J. McEliece and T. H. Palmatier, "Estimating the Size of Huffman Code Preambles," *TDA Progress Reports* 42-114, April-June 1993, pp. 90-95, August 15, 1993.

[4] M. Rabbani and P. W. Jones, *Digital Image Compression Techniques*, Bellingham, Washington: SPIE Press, 1991.

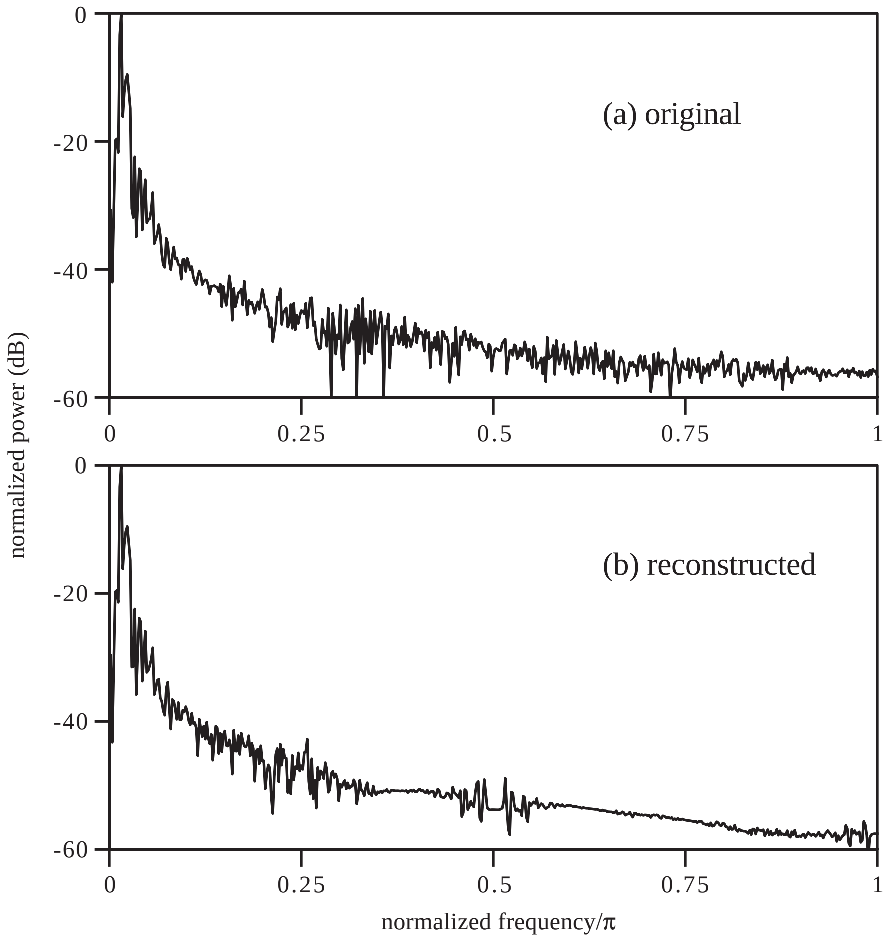

Figure 9: Periodograms of 1024 point BHZ (20 samples/second) background (i.e., non-event) data constructed from (a) original and (b) reconstructed waveform with low resolution quantizer.

[5] A. S. Spanias, S. B. Jonsson, and S. D. Stearns, “Transform Methods for Seismic Data Compression,” *IEEE Trans. Geoscience and Remote Sensing*, vol. 29, no. 3, pp. 407-416, May 1991.

[6] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Englewood Cliffs, New Jersey: P T R Prentice-Hall, 1993.

[7] *Workshop on the Use of Data Compression in Seismic Data*, Workshop Proceedings, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, March 2, 1994.